.

NEUROECONOMICS AND REINFORCEMENT LEARNING. AN EXPLORATORY ANALYSIS

CHAVAGLIA, José (BR), FILIPE, José António (PT), FERREIRA, Manuel Alberto M. (PT)

Abstract. In an era in which the study of the brain gets increasingly featured as an analytical tool of economic phenomena, the understanding of the way Neuroeconomics works as a tool for managers, students and other economic agents is now particularly important. There is now a need to have research that allow the clarification of the analytical decision process and the potential of brain study for decision making process of economic agents. Considering this new reality, a study is presented concerning one of the most important issues for the decision maker, which is the reinforcement learning process.

Key words. Reinforcement learning, Neuroeconomics, Individual behaviour.

Mathematics Subject Classification: 91B06.

1 Introduction

For a long time, philosophers and scholars of human behavior believed (and generally considered) that the individual was able to take his decisions in a rational way and thus it was possible to optimize his performance considering the different decision moments along each day. Maybe the phrase that best represents this is that of the Chilean poet, Pablo Neruda: “you are free to make your own choices but you are a prisoner of the resulting consequences”.

With the emergence of the Neurosciences and their effective research techniques of the brain, the study of human behavior has evidenced a considerable bias about human rationality hypothesis, in particular, in times of economic decision.

It is possible to identify notable scholars of human behavior and neurosciences making references to a set of relating areas in the development of this subject. It is important to consider authors who, among others, are already references nowadays, such as Daniel Kahneman (1934-), António Damásio (1944-), Patrick Renvoisé (2009), Geoffrey Miller (1965-), Nassin Taleb (1960-).

The decision-making problems concerning the “reinforcement learning” have been shown as a kind of infinite possibilities for new studies about the decision-making process from a neuroeconomics point of view (Chavaglia, 2015, p. 99). This article represents another small contribution in this field of study.

The real power is definitely not on the side of whoever is choosing. The decision occurs in a fraction of 2.5 thousandths of a second and several brain mechanisms are involved in the process of creating the preference after occurring the choice. This generates a series

of biases as it is the case of the decision-making process concerning time (Varian, 2006: 595), anchoring (Ariely, 2008: 23; Chavaglia, Filipe and Ramalheiro, 2011: 184), equity and corruption (Akerlof and Shiller, 2010: 11), law of small numbers (Kahneman, 2012: 152), among other deleterious effects of reason.

Therefore, this study aims to provide a contribution considering the way agents realize the neuroeconomics as analytical tool based on a mathematical presentation of the process of "reinforcement learning". However, first of all this requires a formal presentation of Neuroeconomics. For that the article "Neuroeconomics: Decisions in Extreme Situations" (Chavaglia et al., 2015) will be used as a basis.

2 Some Conceptual Considerations

Neuroeconomics is the fusion not just of Neuroscience and Economics as the name directly suggests, but happens also from the junction of many other disciplines (biology, physics, chemistry, statistics, mathematics, psychology, pharmacology, among others). This connection results in a decision-making process that is more 'realistic' and suitable for the everyday moments of life from economic agents' point of view. Neuroeconomics arose from the need of achieving the most reliable results on the individuals economic decisions.

Although considering Neuroeconomics concepts as the "theoretical framework" it is not possible to assert that there are Neuroeconomic theories as it occurs in traditional Economics. Neuroeconomics is the outcome of a set of biological and mathematical results of situations of cerebral processes of people's decision making in Economics. Neuroeconomics is a new field of study in the Economics field that analyzes the relationship between the internal organization of the brain and individuals behavior. It is based on individual decision-making, social interaction and on institutions such as the market (Sandroni, 2007: 907). It is also an emerging transdisciplinary field which uses neuroscience measurement techniques to identify the neural substrates associated to economic decisions (Zak, 2004: 1737).

However, there are some cases in which the combination of concepts and practices of Neuroeconomics with traditional methodologies occurs, as it happens, for example, with the use of applied mathematical axioms (Caplin and Dean, 2007: 16).

This new field of study has within its core the premise that human beings are under bounded rationality and are driven by cognitive biases that are unconsciously derived; for this, Neuroeconomics proposes a particular vision vis-à-vis the traditional vision of traditional Economics. This new vision guarantees a significant importance in the development of economic studies which represent a reliable way for the decision making and for understanding the current complex economic problems in this new era of complex decision making processes.

3 The Study

The study consists in understanding the decision-making process in the economy, in particular with regard to the learning process. This theme should be constant in the study of Neuroeconomics. This is due to the fact the interaction with economic life be constant in the life of an individual. The mode as people create buying habits, negotiation

processes, team management, sales or leadership, for example works as some ways Neuroeconomics deal with, being directly related to the form and type of learning.

3.1 Reinforcement learning in Neuroeconomics

A point that is relevant to the understanding of animal behavior (particularly of human beings) is the stimulus through rewards and punishments. To this end, the study related to the reinforcement learning seems relevant to understand this process.

In the field of reinforcement learning the presence of two approaches is evidenced: Pavlovian conditioning and instrumental conditioning (see Tassi, 2011, p. 30). The first approach got its name because of its creator, Ivan Petrovich Pavlov (1849-1936), winner of the Nobel Prize for Medicine and Physiology in 1904. Pavlov found that some behavioral responses are no-conditioned reflexes, are innate rather than learned, while others are conditioned reflexes. These ones are learned by pairing with pleasant or adverse situations (see Edward, 2009, p. 59). Considering the instrumental process, an agent has control over future stimuli through his actions. It is the case of an animal that determines the release of food by pressing a lever. For Neuroscience still exists the issue of neural implementation of this learning, whereas neural structures and the question of how the information about the environment is stored, and how, from this information, adapted decisions are generated (see Tassi, 2011, p. 30).

This way, it is interesting to note the importance of mathematical modeling of reinforcement learning (see Tassi, 2011, p. 33).

$$V_{new} = V_{previous} + \eta(results - expectations) \quad (1),$$

or

$$V_{new} = V_{previous} + \eta(R - V_{previous}) \quad (2),$$

This means that in an attempt the predictive power of stimulus (V_{new}) will change, for more or for less, just if the value of the result, i.e. the obtained enhancement (R), is greater than or less than the one expected or predicted ($V_{previous}$). If it is the same, no change will occur in the value. The intensity of change will be given for learning coefficient η ($0<\eta\leq1$). The higher the value of learning coefficient η, the greater the importance of the result of the last attempt in determining the value of the stimulus (see Tassi, 2011, p. 33).

Consider now the alternative model for Pavlovian conditioning: the rule of Temporal Difference (TD).

TD model, unlike the previous one, does not deal with time as an object mathematically discrete, but takes into account the temporal relationships between the events in each attempt. Also in TD model, the agent learning goal is different. The agent seeks to make an estimate of the values of the different states or situations, in terms of reinforcements or future total punishments to predict in each state. The previous model, as we have seen, concerns only to learning considering one attempt (see Tassi, 2011, p. 34).

Unlike the previous model, TD considers the effect time for analysis. Another difference refers to the goal of learning. The TD intends to make an estimate of the values of

different states, considering a stimulus for future reward or punishment, as results from the mathematical model. The value of the state "s" at time "t" is:

$$V(s,t) = E[y^0 r_t + y^1 r_{t+1} + y^2 r_{t+2} + y^3 r_{t+3} + \cdots + y^n r_{t+n}] \quad (3),$$

being $0 < y \leq 1$.
The operator $E[.]$ indicates that V_t is the mathematical expectation or average of values in brackets for the various series of previous attempts (the model considers the assumption that the agent followed several times the sequence of states of t = 0 to t = n). The equation (3), considered in its extended form, shows that the value of the state "s" at the time "t" is equal to the value of reinforcement immediately available, summed to the discounted value by y of the reinforcement of the following state, plus the value of reinforcement of the successive state being y^2 discounted, and so on, so that at each temporal step the value of the correspondent reinforcement has a similar devaluation, making that the most distant reinforcements have less value than the latest ones. This is the so called exponential discount. Equation (3) can be written as in the short form below. In order to simplify the notation, V_t instead of $V_{(s,t)}$. will be used.

$$V_t = r_t + yV_{t+1} \quad (4).$$

In this way the learning on TD is identified by the frequent repetition of the succession of states and by repeated estimates correction. Specifying, there is an adjustment made on the difference between the expected value and the value found to update the value of the state (see Tassi, 2011, p. 33).

$$Difference = updated\ state\ of\ reinforcement + y * forecast\ for\ next\ state - forecast\ for\ present\ state:$$

$$\delta_t = r_t + yV_{t+1} - V_t \quad (5).$$

The equation 5 presents the error "δ_t" which is used to update the value "V_t". This equation represents the difference between the expected (V_t) and the obtained one with the immediate reinforcement as the one resulting from the transition to a new state ($r_t + yV_{t+1}$). Obviously δ_t encodes the surprise, the forecast error or the expectation's violation. It will be used for updating the value of the state after the multiplication by a constant which determines, as in the previous case, how much the value will be changed depending on this signal.

$$V_{new} = V_{previous} + \eta\, \delta_t \quad (6).$$

In this case, updating will depend on the learning coefficient η (η ≤ 1 0 <), and establishes how much the result of the recent experience will be crucial (see Tassi, 2011, p. 37).
Let's see now the operating conditioning. In this model the transition between states will depend on the actions of the agent and the agent's goal becomes choosing to be

conditioned according to states associated with the largest sum of present and future reinforcements. Every action determines the transition and immediate reinforcement, but also the possible subsequent transitions and consistent reinforcements. This way, when the agent receives a reinforcement after a sequence of several actions it is necessary to find out what is the previous action that has increased the probability of reinforcement. Such a problem is called "temporal credit allocation" (see Tassi, 2011, p. 37).
Considering the mathematical formulation,

$$p(action|state)_{new} = p(action|state)_{previous} + \eta\delta \quad (7).$$

Another reinforcement learning algorithm to model the operating conditioning is "Q-learning". In this algorithm, the agent learns directly to find the best probability of action in every state without having to learn the value of each state (see Tassi, 2011, p. 38).

$$q(s,a) = E[r_t + y^1 r_{t+1} + y^2 r_{t+2} + \cdots + y^n r_{t+n} \,|state(t) = s, action(t) = a] \quad (8).$$

This equation, where $Q(s,a)$ is the mean value of the action (a) in the state (s), gives how much the agent can, on average, expect if in the state s he chooses the action a and then choosing always the action that previous experience taught be the highest return. By making this choice, the agent moves to the next state and receives or not a reinforcement. At this point, the difference between the forecast and the true value will be calculated, δ_t, and this will be used to update the value of that state-action,

$$Q(s,a): Q(s,a)_{new} = Q(s,a)_{previous} + \eta\delta_t \quad (9),$$

while other values of Q for the other state and action pairs remain unchanged. The process Q-learning consists on the repetition of three steps: (1) making the prediction of the expected reinforcement of candidate actions in that state; (2) selecting the action with the highest human reinforcement; and (3) updating the Q value for that chosen pair state-action, by using the error or discrepancy between the obtained and the expected value (see Tassi, 2011, p. 39).

4 Conclusions

The limited capacity of decision-making by individuals strongly suggests that the individual must support his analysis on models which are closer to the reality of the decision maker. Considering that, neuroeconomics models are a real possibility.
However, it is necessary to use more advanced research techniques to study the brain of decision makers. Only by doing this, it will be possible to verify how the endogenous mechanisms of brain work at the same time agents make their choices. Certainly the qualitative gains in the analysis result very considerable.

In addition to the use of more precise research forms, it is also important to submit this topic to a greater number of researches so that a major number of hypotheses will be tested in decision-making situations, particularly, in the area of economic decisions.

Nevertheless, it is necessary to be prudent about the generalization of results considering the scientific and practical uses of these results. Neuroeconomics is still considered a new field of knowledge. The contribution of neuroscience and Economics considered together emerges as the formation of this new field. However, the results point to the fact that neuroeconomics appears itself as a completely new field because its premises, tests, applications and discussions have advanced to a single direction and distinct from paths that neuroscience followed and completely different from the path traced by the Orthodox economic science, if not often antagonistic.

Specifically considering the issue of reinforcement learning, it is clear that its use - for the conceptual and practical understanding in daily economic institutions - has gains in analytical quality to address this issue on the prism of Neuroeconomics. It allows a systematically view of how learning triggers human behaviour. Subsequently, the mathematical model of the types of learning contemplates this understanding in a little more scientific way to determine quantitative interactions. This appreciation immediately facilitates the study and dissemination in scientific media, which in turn, help to consolidate the theme as a field of study.

Plainly the results found in many studies of neuro and behavioral economics show that humans are not fully rational in their decisions, for instance during elections on when deciding about other situations of complete uncertainty. This fact also emerges from data results obtained and worked on this analysis.

References

[1] AINSLIE, G., MOTEROSSO, J. (2004), A market place in the brain?, *Science*, Vol. 306, No. 5695, pp. 421- 423.

[2] AKERLOF, G., SCHILLER, R. (2010), O *espírito animal*, Rio de Janeiro, Campus.

[3] ARIELY, D. (2008), *Previsivelmente irracional*, Rio de Janeiro, Campus.

[4] BREMMER I. (2010), *O fim do livre mercado*, São Paulo, Saraiva.

[5] BROEKHOFFB, M. (2014), Emotions in television: crucial for the customer experience, *Neuromarketing: theory & practice*, 9, 4-5.

[6] CALEIRO, A. (2013a), A Self-Organizing Map of the Elections in Portugal, *The IIOAB Journal*, Special Issue (Neuroscience in Economic Decision Making), 4: 3, April-June, 9-14.

[7] CALEIRO, A. (2013b), How to Classify a Government: Can a perceptron do it?, *International Journal of Latest Trends in Finance and Economic Sciences*, 3: 3, September, 523-529.

[8] CAPLIN, A., DEAN, M. (2007), Axiomatic neuroeconomics, neoclassical economic approach, *Neuroeconomics: decision making and the brain*, Elsevier.

[9] CARVALHO, J. E. (2009), *Neuroeconomia*, Lisboa, Sílabo.

[10] CHAVAGLIA, J. N. (2014), *O Sector Elétrico Brasileiro à Luz da Neuroeconomia: O Caso das Energias Renováveis*, PhD Thesis, ISCTE-IUL. Lisboa.

[11] CHAVAGLIA, J. N., FILIPE, J. A. and RAMALHEIRO, B. (2013), Neuroeconomics: the effect of context in decisions relating to the Brazilian electric sector, *IIOABJ;* Vol. 4; Issue 3; 2013: 38-44.

[12] CHAVAGLIA, J. N., FILIPE, J. A. and RAMALHEIRO, B. (2011), Neuromarketing: Consumers and the Anchoring Effect. *International Journal of Latest Trends in Finance and Economic Sciences*, Vol. 1(4), pp. 183-189.

[13] GRAEFF, F. (2003) Serotonin, the periaqueductal gray and panic, *Neuroscience and Biobehavioral Reviews*, 28, 239-259.

[14] GREENE, J. D., NYSTROM, L. E., NYSTROM, A D., ENGEL, A D., DARLEY, J. M. and COHEN, J. D. (2004), The neural bases of cognitive conflict and control in moral judgement. *Neuron*, 44, pp389-400.

[15] GREENE, J. D., SOMMERVILLE, R. B., NYSTROM, L. E., DARLEY, J. M. and COHEN, J. D. (2001), An fMRI investigation of emotional engagement in moral judgment. *Science*, 293, pp 2105-2108.

[16] KAHNEMAN, D. (2012), *Rápido e devagar: duas formas de pensar*, Rio de Janeiro, Objetiva.

[17] KAHNEMAN, D., KITSCH, J. L., THALER, R. (1990), Experimental tests of the endowment effect and the Coase theorem, *Journal of Political Economy*, 98, 1325-1348.

[18] KAHNEMAN, D., TVERSKY, A. (1974), *Judgment under uncertainty: heuristics and biases***,** Science.

[19] KlARIK, J. (2012), *Estamos Cegos*, São Paulo, Planeta.

[20] MARTIN, N. (2013), Proving a New Discipline, *Neuromarketing: theory & practice*, 7, 7.

[21] NEUMAERKER, B. (2007), Neuroeconomics and the Economic Logic of Behavior, *Analyse & Kritik*, 29, 60-85.

[22] ROCHA, A. F.; MASSAD, E.; ROCHA, F. T., The Neuroeconomics of Emotional Conflicts in Moral Dilemma Judgment. Available in http://www.eina.com.br/trabalhos/dilema.pdf, assessed on April 9, 2013.

[23] SANDRONI, P. (2007), *Novíssimo dicionário de economia*, São Paulo, Best Seller

[24] TASSI, L. E. (2011), *Desempenho de Ratos em jogo estratégico e modulação dopaminérgica*, Tese de Doutorado, Universidade de São Paulo, Departamento de Fisiologia.

[25] TREPEL, C.; FOX, C. R.; POLDALOCK, R. A. (2005), Prospect theory on the brain? Toward a cognitive neuroscience of decision under risk, *Cognitive Brain Research*, 23: 1, 34-50.

[26] Varian, H. (2006), *Microeconomia: princípios básicos*, Rio de Janeiro, Campus.

[27] Zack, P. J. (2004), *Neuroeconomics*, The royal society, 359, 1737-1748.

Current addresses

José Chavaglia Neto
INSTITUTO UNIVERSITÁRIO DE LISBOA (ISCTE-IUL)
AV. DAS FORÇAS ARMADAS
1649-026 LISBOA, PORTUGAL
TELEFONE: + 351 21 790 37 03
FAX: + 351 21 790 39 41
E-MAIL: jnchavaglia@gmail.com

José António Candeias Bonito Filipe, Professor Auxuliar com Agregação
INSTITUTO UNIVERSITÁRIO DE LISBOA (ISCTE-IUL)
BRU – IUL, ISTAR-IUL
AV. DAS FORÇAS ARMADAS
1649-026 LISBOA, PORTUGAL
TELEFONE: + 351 21 790 37 03
FAX: + 351 21 790 39 41
E-MAIL: josé.filipe@iscte.pt

Manuel Alberto M. Ferreira, Professor Catedrático
INSTITUTO UNIVERSITÁRIO DE LISBOA (ISCTE-IUL)
BRU – IUL, ISTAR-IUL
AV. DAS FORÇAS ARMADAS
1649-026 LISBOA, PORTUGAL
TELEFONE: + 351 21 790 37 03
FAX: + 351 21 790 39 41
E-MAIL: manuel.ferreira@iscte.pt